A Societal Sentiment Analysis: Predicting the Values and Ethics of Individuals by Analysing Social Media Content

Tushar Maheshwari[1*] **Aishwarya N Reganti**[1*] **Samiksha Gupta**[2] **Anupam Jamatia**[3]
Upendra Kumar[1]   **Björn Gambäck**[4,5]   **Amitava Das**[1]

[1] Indian Institute of Information Technology, Sri City, Andhra Pradesh, India
[2] Indian Institute of Technology, Varanasi, Uttar Pradesh, India
[3] National Institute of Technology, Agartala, Tripura, India
[4] Norwegian University of Science and Technology, Trondheim, Norway
[5] SICS Swedish ICT AB, Kista, Sweden

[1] `{tushar.m14,aishwarya.r14,upendra.k14,amitava.das}@iiits.in`
[2] `samiksha.gupta.apm13@itbhu.ac.in`
[3] `anupamjamatia@gmail.com`   [4] `gamback@idi.ntnu.no`

Abstract

To find out how users' social media behaviour and language are related to their ethical practices, the paper investigates applying Schwartz' psycholinguistic model of societal sentiment to social media text. The analysis is based on corpora collected from user essays as well as social media (Facebook and Twitter). Several experiments were carried out on the corpora to classify the ethical values of users, incorporating Linguistic Inquiry Word Count analysis, n-grams, topic models, psycholinguistic lexica, speech-acts, and non-linguistic information, while applying a range of machine learners (Support Vector Machines, Logistic Regression, and Random Forests) to identify the best linguistic and non-linguistic features for automatic classification of values and ethics.

1 Introduction

In the recent years, there have been significant efforts on determining the opinion/sentiment/emotion about a specific topic held by the author of a piece of text, and on automatic sentiment strength analysis of text, classifying it into either one of the classes positive, negative or neutral, or into Ekman's classes of happy, sad, anger, fear, surprise, and disgust. However, the intrinsic value of the lives we lead reflects the strength of our values and ethics which guide our social practices, attitude and behaviour. This paper reports work on investigating

a psycholinguistic model, the Schwartz model (Schwartz and Bilsky, 1990; Schwartz, 2012), and applying it to social media text. It will here be referred to as a societal sentiment model, since societal values grow from the interactions, and the views and sentiment of the society are key to ethical practices. No computational model for Schwartz' Values has been tested or examined before as such, but there has been a growing interest in the scientific community on doing automatic personality recognition, commonly using the Big 5 factor model (Goldberg, 1990) that defines personality traits such as openness, conscientiousness, extraversion, agreeableness, and neuroticism.

The Schwartz values model defines ten distinct ethical values (henceforth only values), that respectively are: ***Achievement*** sets goals and achieves them; ***Benevolence*** seeks to help others and provide general welfare; ***Conformity*** obeys clear rules, laws and structures; ***Hedonism*** seeks pleasure and enjoyment; ***Power*** controls and dominates others, controls resources; ***Security*** seeks health and safety; ***Self-direction*** wants to be free and independent; ***Stimulation*** seeks excitement and thrills; ***Tradition*** does things blindly because they are customary; ***Universalism*** seeks peace, social justice and tolerance for all.

Deeper understanding of human beliefs, attitudes, ethics, and values has been a key research agenda in Psychology and Social Science research for several decades. One of the most accepted and widely used frameworks is Schwartz 10-Values model, has seen great success in psychological research as well as in other fields. The ten basic values are related to various outcomes and ef-

---

* The two first authors contributed equally.

fects of a person's role in a society (Argandoña, 2003; Agle and Caldwell, 1999; Hofstede et al., 1991; Rokeach, 1973). Schwartz values have also proved to provide an important and powerful explanation of consumer behaviour and how they influence it (Kahle et al., 1986; Clawson and Vinson, 1978). Moreover, there are results that indicate how values of workforce and ethical practice in organisations are directly related to transformational and transactional leadership (Hood, 2003).

We believe that these kind of models may become extremely useful in the future for various purposes like Internet advertising (specifically social media advertising), community detection, computational psychology, recommendation systems, sociological analysis (for example East vs West cultural analysis) over social media.

In order to experiment with this, three corpora have been collected and annotated with Schwartz values. Two of the corpora come from popular social media platforms, Facebook and Twitter, while the third corpus consists of essays. A range of machine learning techniques has then been utilized to classify an individual's ethical practices into Schwartz' classes by analyzing the user's language usage and behaviour in social media. In addition to identifying the ten basic values, Schwartz' theory also explains how the values are interconnected and influence each other, since the pursuit of any of the values results in either an accordance with one another (e.g., Conformity and Security) or a conflict with at least one other value (e.g., Benevolence and Power). The borders between the motivators are artificial and one value flows into another. Such overlapping and fuzzy borders between values make the computational classification problem more challenging.

The paper is organized as follows. Section 2 introduces related work in the area. Details of the corpora collection and annotation are given in Section 3. Section 4 reports various experiments on automatic value detection, while Section 5 discusses the performance of the psycholinguistic experiments and mentions possible future directions.

2   Related Work

State-of-the-art sentiment analysis (SA) systems look at a fragment of text in isolation. However, in order to design a Schwartz model classifier, we require a psycholinguistic analysis. Therefore, textual features and techniques proposed and dis-

cussed for SA are quite different from our current research needs. Hence, we will here focus only on previous research efforts in automatic personality analysis that closely relate to our research work. Personality models can be seen as an augmentation to the basic definition of SA, where the aim is to understand sentiment/personality at person level rather than only at message level.

In recent years, there has been a lot of research on automated identification of various personality traits of an individual from their language usage and behaviour in social media. A milestone in this area was the 2013 Workshop and Shared Task on Computational Personality Recognition (Celli et al., 2013), repeated in 2014 (Celli et al., 2014). Two corpora were released for the 2013 task. One was a Facebook corpus, consisting of about 10,000 Facebook status updates of 250 users, plus their Facebook network properties, labelled with personality traits. The other corpus comprised 2,400 essays written by several participants labelled with the personalities. Eight teams participated in the shared task. The highest result was achieved by Markovikj et al. (2013) with an F-score of 0.73 (average for all the traits). The main methods and features (linguistic as well as non-linguistic) used by the participant groups were as follows.

Linguistic Features: The participating teams tested several linguistic features. Since n-grams are known to be useful for any kind of textual classification, all the teams tested various lengths of n-grams (uni, bi, and tri-grams). Categorical features like part-of-speech (POS), word level features like capital letters, repeated words were also used. Linguistic Inquiry Word Count (LIWC) features were used by all the teams as their baselines. LIWC (Pennebaker et al., 2015) is a hand-crafted lexicon specifically designed for psycholinguistic experiments. Another psycholinguistic lexicon called MRC (Wilson, 1988) was also used by a few teams, as well as lexica such as SentiWordNet (Baccianella et al., 2010) and WordNet Affect (Strapparava and Valitutti, 2004). Two more important textual features were discussed by the participating teams. Linguistic nuances, introduced by Tomlinson et al. (2013), is the depth of the verbs in the Wordnet troponymy hierarchy. Speech act features were utilized by Appling et al. (2013): the authors manually annotated the given Facebook corpus with speech acts and reported their correlation with the personality traits.

732

Here we briefly describe some people. Please read each description and think about how much each person is or is not like you. Tick the box to the right that shows how much the person in the description is like you.

	HOW MUCH LIKE YOU IS THIS PERSON?					
	Very much like me	Like me	Some-what like me	A little like me	Not like me	Not like me at all
1. Thinking up new ideas and being creative is important to her. She likes to do things in her original way. **SD**	6	5	4	3	2	1
2. It is important to her to be rich. She wants to have a lot of money and expensive things. **PO**	6	5	4	3	2	1
3. She thinks it is important that every person in the world be treated equally. She believes everyone should have equal opportunities in life. **UN**	6	5	4	3	2	1
4. Its important to her to show her abilities. She wants people to admire what she does. **AC**	6	5	4	3	2	1
5. It is important to her to live in secure surroundings. She avoids anything that might endanger her safety. **SE**	6	5	4	3	2	1

Table 1: Instructions and format of the Portrait Values Questionnaire. For each statement, the respondents should answer the question "*How much like you is this person?*" by checking one of the six boxes.

Non-Linguistic Features: All teams used Facebook network properties including network size, betweenness centrality, density and transitivity, provided as a part of the released dataset.

3 Corpus Acquisition

To start with, we ask a very fundamental question: *whether social media is a good proxy of the original (real life) society or not.* Back et al. (2010) and Golbeck et al. (2011) provide empirical answers to this question. Their results respectively indicate that, in general, people do not use virtual desired/bluffed social media profiles to promote an idealized-virtual-identity and that a user's personality can be predicted from his/her social media profile. This does not mean that there are no outliers, but our corpus collection was grounded on the assumption that it is true for a major portion of the population that social media behaviour to a large extent mirror that of the actual human society. Two of the most popular social media platforms, Twitter and Facebook, were chosen as sources for the corpora to validate this assumption. In addition, an essay corpus was collected. These three diverse corpora were then used for training and testing Schwartz values analysis methods.

3.1 Questionnaire for Self-Assessment

A standard method of psychological data collection is through self-assessment tests, popularly known as psychometric tests. In our experiments, self-assessments were obtained using male/female versions of PVQ, the Portrait Values Questionnaire (Schwartz et al., 2001). The participants were asked to answer each question on a 1–6 Likert rating scale.[1] A rating of 1 means "*not like me at all*" and 6 means "*very much like me*". An example question is "*He likes to take risks. He is always looking for adventures.*" where the user should answer while putting himself in the shoes of "He" in the question. A few exemplary items as well as the instructions and format of the written form of the PVQ are presented in Table 1.

The standard practice is to ask a fixed number of questions per psychological dimension. Here there are five questions for each of the ten Values classes, resulting in a 50 item questionnaire. Once all the questions in the PVQ have been answered, for each user and for each Values class, a score is generated by averaging all the scores (i.e., user responses) corresponding to the questions in that class, as described by Schwartz (2012). Further, the rescaling strategy proposed by Schwartz (2012) was used to eliminate randomness from each response given by a user as follows: For each user, the mean response score was first calculated considering all the responses s/he provided, and then the mean score from each response was subtracted. See Schwartz (2012) for more details on PVQ and the score computation mechanism.

The ranges of scores obtained from the previous rescaling method may vary across different Values classes. For instance, the ranges of the rescaled scores for the Essay corpus are as follows: Achievement $[-4.12, 3.36]$, Benevolence $[-1.56, 3.39]$, Conformity $[-3.35, 3.01]$, Hedo-

---

[1] http://www.simplypsychology.org/likert-scale.html

733

nism $[-5.18, 4.35]$, Power $[-6.0, 2.27]$, Security $[-2.60, 2.40]$, Self-Direction $[-1.61, 3.40]$, Stimulation $[-5.0, 2.63]$, Tradition $[-4.49, 3.35]$, and Universalism $[-3.33, 3.30]$.[2]

Hence the standard normalisation formula was applied to move the ranges of the different Values classes to the $[-1, 1]$ interval:

$$x_{scaled} = \frac{2 * (x - x_{min})}{x_{max} - x_{min}} - 1$$

A 'Yes' or 'No' binary value was assigned to each Values class: if the score was less than 0, the class was considered to be negative, indicating absence of that Values trait for the particular user; while scores ≥ 0 were considered to be positive, indicating the presence of that trait for the user. We will use the real scores ranging $[-1, 1]$ for the regression experiments mentioned in Section 4.

Reports of psychological analysis always depend on how the target population is chosen. Therefore while we are hypothesising that a few people are more Power oriented, an open question that remains unanswered is whom they are more Power oriented than. For example, if we (hypothetically) choose parliamentarians / politicians as participants in an experiment, then the entire examined population will likely turn out to be Power oriented. Therefore, it makes sense to normalise the obtained data into two groups $[-1, 0)$ and $[0, 1]$ and proclaim that people with $[0, 1]$ range scores are relatively more Power (or any other Value) oriented than the people having score ranging $[-1, 0)$.

The same normalisation mechanism was applied to all the corpora, but also after normalisation the different Values distributions were imbalanced (with the Facebook data being the most imbalanced). One possible reason behind such imbalanced distributions is that the portion of the real population using social media is slightly biased towards some Values types due to several societal reasons such as educational/family background, age group, occupation, etc. Another reason could be that the divisions between different value types simply never are balanced in any population. However, analysing such societal traits is a separate research direction altogether and out of the scope of the current study.

The PVQ questionnaire setting described above was used to separately collect textual user data separately for the Essay, Facebook, and Twitter corpora, as discussed in the rest of this section.

3.2 Essay Corpus

The Essay corpus was collected using the Amazon Mechanical Turk (AMT)[3] crowd-sourcing service. The turkers (users providing responses on AMT) were asked to compose an essay on the most important values and ethics guiding their lives, and to answer the PVQ questionnaire. A total of 981 users participated in the essay writing. However, not all the responses were useful for the analysis, since some participants did not answer all the questions and some did not write the essay carefully. For example, one user wrote: "*I don't really have a guide in life. I go by what sounds and feels good. that means what makes me happy rather that effects others or not.*" Filtering out such users, data from 767 respondents was retained.

3.3 Twitter Corpus

In the first quarter of 2016, the micro blogging service Twitter averaged 310 million monthly active users,[4] with around 6,000 tweets being posted every second. Therefore, Twitter came as the second natural choice as data source. The data collection was crowd-sourced using Amazon Mechanical Turk, while ensuring that the participants came from various cultures and ethnic backgrounds: the participants were equally distributed, and consisted of Americans (Caucasian, Latino, African-American), Indians (East, West, North, South), and a few East-Asians (Singaporeans, Malaysian, Japanese, Chinese). The selected Asians were checked to be mostly English speaking.

The participants were requested to answer the PVQ questionnaire and to provide their Twitter IDs, so that their tweets could be crawled. However, several challenges have to be addressed when working with Twitter, and a number of iterations, human interventions and personal communications were necessary. For example, several users had protected Twitter accounts, so that their tweets were not accessible when using the Twitter API. In addition, many users had to be discarded since they had published less than 100 tweets, making

---

[2]The distribution of value types over a corpus was analysed using the Bienaymé-Chebyshev Inequality (Bienaymé, 1853; Tchébichef, 1867), showing that, e.g, most Achievement instances (89%) were in the range $[-2.96, 2.84]$.

[3]https://www.mturk.com/
[4]statista.com/statistics/282087/
number-of-monthly-active-twitter-users

Corpus	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN	Avg
Essay	65.70	52.41	63.10	54.40	40.40	52.50	54.00	52.50	43.20	54.43	66.10
Twitter	81.00	78.70	73.30	77.10	50.10	76.30	83.40	73.60	52.60	82.00	72.80
Facebook	88.13	93.22	91.52	86.44	46.67	98.30	89.83	86.44	71.18	94.91	84.67

Table 2: Flat distribution of Schwartz' value types in the corpora: Achievement (AC), Benevolence (BE), Conformity (CO), Hedonism (H), Power (PO), Security (SE), Self-Direction (SD), Stimulation (ST), Tradition (TR), Universalism (UN). The last column gives the Average Majority Baselines.

Personality	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
Achievement (AC)	—	28.31	19.49	29.41	41.54	15.81	11.77	19.85	41.91	17.28
Benevolence (BE)	24.12	—	19.84	31.12	52.92	18.68	10.51	22.18	42.80	7.00
Conformity (CO)	18.59	23.42	—	35.32	47.58	12.64	17.47	24.91	35.32	15.99
Hedonism (HE)	17.60	24.04	25.32	—	43.35	21.03	9.01	14.60	45.92	12.88
Power (PO)	12.64	33.52	22.53	27.47	—	17.58	13.74	17.03	41.21	20.33
Security (SE)	17.63	24.82	15.47	33.81	46.04	—	13.31	21.94	38.49	14.39
Self-Direction (SD)	21.05	24.34	26.97	30.26	48.35	20.72	—	20.72	47.04	12.50
Stimulation (ST)	18.66	25.37	24.63	25.75	43.66	19.03	10.08	—	42.91	16.04
Tradition (TR)	18.13	23.83	9.84	34.72	44.56	11.40	16.58	20.73	—	17.10
Universalism (UN)	24.75	20.07	24.41	32.11	51.51	20.40	11.04	24.75	46.49	—

Table 3: Fuzzy distributions of Schwartz' ten personality value types in the Twitter corpus.

them uninteresting for statistical analysis. In addition, some extreme cases when users mentioned someone else's (some celebrity's) Twitter account, had to be discarded. The open source free Twitter API: Twitter4J[5] also has a limit of accessing only the current 3,200 tweets from any user. To resolve this issue, an open source Java application (Henrique, 2015) was used.

At the end of the data collection process, data from 367 unique users had been gathered. The highest number of tweets for one user was 15K, while the lowest number of tweets for a user was a mere 100; the average number of messages per user in the Twitter corpus was found to be 1,608.

3.4 Facebook Corpus

Facebook (FB) is the most popular social networking site in the world, with 1.65 billion monthly active users during the first quarter of 2016.[6] Therefore, Facebook was a natural first choice for corpus collection, but since the privacy policy of Facebook is very stringent, accessing Facebook data is challenging. To collect the corpus, a Facebook Canvas web-application was developed using Facebook Graph API and Facebook SDK v5 for PHP library. Undergraduate students of two Indian institutes (NIT, Agartala, Tripura and IIIT, Sri City, Andhra Pradesh) were contacted for the data

collection. The application was circulated among the students and they were requested to take part in the PVQ survey and to donate their FB Timeline data and friend list data. Timeline data includes their own posts and all the posts they are tagged in, and posts other people made on their Timeline.

So far, data from 114 unique users has been collected, but the data is highly imbalanced (for some value types the distributions of 'Yes' and 'No' classes were in 90:10 ratio). Crowd-sourcing is a cheap and fast way to collect data, but unfortunately some annotators chose random labels to minimize their cognitive thinking load. These annotators can be considered as spammers and make aggregation of crowd-sourced data a challenging problem, as discussed in detail by Hovy et al. (2013). To filter out spammers, the MACE (Hovy et al., 2013) tool was used and data from 54 users discarded, so the final dataset includes only 60 participants. The average number of messages per user in the Facebook corpus is 681.

3.5 Corpus Statistics

Categorical flat distributions are reported in Table 2. Schwartz' model defines fuzzy membership, which means that anyone having a Power orientation can have the Achievement orientation as well. To understand this notion vividly, we have reported the fuzzy membership statistics from the Twitter data in Table 3 (due to space limitations, statistics for the other corpora are not reported).

---

[5] http://twitter4j.org/

[6] http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Feature Ablation	AC	BE	CO	HE	PO	SE	SD	ST	TR	UN
Before Ablation	65.84	56.06	64.02	58.02	58.80	53.06	60.89	56.58	64.28	65.58
After Ablation	65.84	58.54	64.80	58.93	59.58	55.80	61.53	56.84	65.06	66.10
Number of features	52	37	65	38	54	47	65	53	39	48

Table 4: Best LIWC feature selection (by accuracy) for each of Schwartz' ten personality value types. The values in the 'Before Feature Ablation' row are based on the full feature set (69 features).

The statistics in Table 3 show how the ten values are interconnected and influence each other, supporting the basic assumption of the Schwartz model that the borders between value classes are artificial and that one value flows into the next.

4 Automatic Identification Experiments

Several experiments were performed to get a better understanding of the most appropriate linguistic and non-linguistic features for the problem domain. The experiments were designed as a single label classification task (each input corresponds to one target label) with 20 classes, with 'Yes' and 'No' classes for each of the ten Schwartz values. Ten different classifiers were trained, each for a particular value type. Each classifier predicts whether the person concerned is positively or negatively inclined towards the given Schwartz value.

The versions implemented in WEKA (Witten and Frank, 2005) of three different machine learning algorithms were used in the experiments: Sequential Minimal Optimization (SMO; a version of Support Vector Machines, SVM), Simple Logistic Regression (LR), and Random Forests (RF). In all the mentioned experiments the corpora were pre-processed, i.e., tokenized by the CMU tokenizer (Gimpel et al., 2011) and stemmed by the Porter Stemmer (Porter, 1980). All the lexica were also stemmed in the same way before usage and all results reported below were obtained using 10-fold cross validation on each of the corpora.

4.1 Linguistic Features

LIWC Analysis: LIWC (Pennebaker et al., 2015) is a well developed hand-crafted lexicon. It has 69 different categories (emotions, psychology, affection, social processes, etc.) and almost 6,000 distinct words. The 69 categorical features were extracted as user-wise categorical word frequencies. As the text length (for the Essay corpus) or number of messages (Twitter and FB corpora) varies from person to person, Z-score normalization (or standardization) was applied using

the equation: $\hat{x} = (x - \mu)/\sigma$, where x is the 'raw frequency count', μ and σ are respectively the mean and standard deviation of a particular feature. After normalizing, each feature vector value is centered around $\sigma = (0, 1)$. This normalization led to an increase in the accuracy figures in many of the cases.

To investigate how each LIWC feature contributes, feature ablation was performed and the Pearson correlations of LIWC features vs value types were analysed. The final classifiers were trained using only the features that were contributing for a particular value type. This resulted in a performance boost and also gave reduced time complexity (both model training and testing times). Table 4 contains detailed categorical features for each value type for the SMO model, and, e.g., shows that the same accuracy (65.84%) for the Achievement class as obtained by using the full 69 feature set also can be obtained by using only 52 LIWC features. Moreover, the lowest obtained accuracy 53.06% for the Security class increased to 55.80% when considering only 47 features. Clearly, the details of which features actually contribute to each class cannot be included here (for space reasons), but the important lesson is that it is possible to reduce the feature set and increase performance in this way.

n-grams: In line with the systems discussed in Section 2, n-gram features were added to the LIWC baseline. In a first run, the top 20% of the most frequent uni-grams from the Essay corpus were included as new features, resulting in a 1452+69 feature set. Unexpectedly, SMO's accuracy dropped by an average of 8.60%. The Achievement and Conformity values suffered the maximum performance drop, whereas Security and Hedonism had a slight increase in accuracy. Random Forests performed well in many of cases, except for the Security and Benevolence classes.

In a second iteration, categorical (value wise) n-grams features were selected and used. The resulting feature set sizes differ for each of the ten

values, ranging from the lowest number 886+69 for Power to the highest 1176+69 for Universalism. Marginally better performance was recorded.

n-grams (word grams) with various sizes of *n*, ranging from 2, 3, 4, 5, and so on, have different impact on performance on different kinds of applications. Commonly, bi-grams are better features for many text classification tasks. So, in a third iteration we tested system performance using bi-grams as added features with LIWC. As the total possible combinations of bi-grams are quite high, only the top 2,000 frequent bi-grams were included, resulting in 2000+69 features. There was no significant performance gain in this experiment on the Essay corpus, so this feature was not tested for the other two corpora.

Topic Modeling: In order to find out the bag-of-words features for each value type, i.e., the vocabulary that a person uses more frequently, the MALLET (McCallum, 2002)[7] topic modelling toolkit was used to extract a number of topics. MALLET uses Gibbs Sampling and Latent Dirichlet Allocation (LDA). In a pre-processing stage, stop words were removed and case was preserved. For the Essay corpus, we tested with different number of topic clusters of sizes 10, 20, 50, 75, and 100, and observed that 50 was the most suitable number. Each of the 50 topics contained an average of 19 words (the topic key words indentified by MALLET), each with a specific weight attached. The top 5 topics were chosen for each value type, according to these weights, and the words of these topics were added as a new feature set along with the LIWC baseline features.

It was also observed that the rankings of the top 5 topics were almost similar for each Schwartz value. The accuracies obtained were almost similar to the accuracies obtained in the previous experiments; however, this time, since the dimension of the feature set is much smaller, the time complexity decreased by almost a factor of 10. Hence the topic modelling was repeated for the social media corpora from Facebook and Twitter, but resulting in a different number of topic clusters, namely 89. Added to the 69 LIWC features this thus resulted in a total of 158 features.

Psycholinguistic Lexica: In addition to the base feature set from LIWC, two other psycholinguistic lexica were added: the Harvard General Inquirer (Stone et al., 1966) and the MRC psycholinguistic database (Wilson, 1988). The Harvard General Inquirer lexicon contains 182 categories, including two large valence categories positive and negative; other psycholinguistic categories such as words of pleasure, pain, virtue and vice; words indicating overstatement and understatement, often reflecting presence or lack of emotional expressiveness, etc. 14 features from the MRC Psycholinguistic lexicon were included, namely, number of letters, phonemes and syllables; Kucera-Francis frequency, number of categories, and number of samples; Thorndike-Lorge frequency; Brown verbal frequency; ratings of Familiarity, Concreteness, Imagability and Age of acquisition; and meaningfulness measures using Colorado Norms and Pavio Norms. In order to get these MRC features a machine readable version of it has been used.[8] Feature ranking was done by evaluating the contribution of each feature in an SMO classifier.

In addition, the sensorial lexicon Sensicon (Tekiroğlu et al., 2014) was used. It contains words with sense association scores for the five basic senses: Sight, Hearing, Taste, Smell, and Touch. For example, when the word 'apple' is uttered, the average human mind will visualize the appearance of an apple, stimulating the eye-sight, feel the smell and taste of the apple, making use of the nose and tongue as senses, respectively. Sensicon provides a numerical mapping which indicates the extent to which each of the five senses is used to perceive a word in the lexicon. Again, feature ablation was performed and the (Pearson) correlations of lexicon features vs values analysed. Finally, classifiers were trained using only contributing features for a particular value.

Speech Act Features: The way people communicate, whether it is verbally, visually, or via text, is indicative of Personality/Values traits. In social media, profile status updates are used by individuals to broadcast their mood and news to their peers. In doing so, individuals utilize various kinds of speech acts that, while primarily communicating their content, also leave traces of their values/ethical dimensions behind. Following the hypothesis of Appling et al. (2013), speech act features were applied in order to classify personalities/values. However, for this experiment the speech act classes were restricted to 11 major categories: Statement Non-Opinion (SNO), Wh

---

[7]http://mallet.cs.umass.edu

[8]http://ota.oucs.ox.ac.uk/headers/1054.xml

Speech Act	SNO	Wh	YN	SO	AD	YA	T	AP	RA	A	O	Avg.
Distribution	33.37	11.45	15.45	5.16	6.88	15.08	0.41	3.26	0.71	0.07	14.59	
F_1-score	0.45	0.88	0.88	0.72	0.45	0.60	0.72	0.60	0.12	0.77	0.12	0.69

Table 5: Speech act class distributions in the corpus (in %) and speech act classifier performance

Question (Wh), Yes-No Question (YN), Statement Opinion (SO), Action Directive (AD), Yes Answers (YA), Thanking (T), Appreciation (AP), Response Acknowledgement (RA), Apology (A) and others (O), hence avoiding having 43 fine-grained speech act classes.[9]

A corpus containing 7K utterances was collected from Facebook and Quora pages, and annotated manually. Motivated by the work by Li et al. (2014), this corpus was used to develop an SVM-based speech act classifier using the following features: bag-of-words (top 20% bigrams), presence of "wh" words, presence of question marks, occurrence of "thanks/thanking" words, POS tags distributions, and sentiment lexica such as the NRC lexicon (Mohammad et al., 2013), SentiWordNet (Baccianella et al., 2010), and WordNet Affect (Strapparava and Valitutti, 2004).

The categorical corpus distribution and the performance of the final classifier are reported in Table 5, showing an average F_1-score of 0.69 in 10-fold cross validation. Automatic speech act classification of social media conversations is a separate research problem, which is out of the scope of the current study. However, although the speech act classifier was not highly accurate in itself, the user specific speech act distributions (in %) could be used as features for the psycholinguistic classifiers (resulting in 11 additional features). Experiment on the Essay and Facebook corpora showed only 1.15% and 1% performance gain, respectively, whereas on the Twitter Corpus, a noticeable performance improvement of 6.12% (F-measure) was obtained. This indicates that speech acts are important signals of psychological behaviour, so even though the speech act classifier performs poorly, the extracted information is relevant.

4.2 Non-Linguistic Features

Social network structure is very useful to predict any person's intrinsic value. For each user in the Twitter corpus, the total number of tweets or messages, total number of likes, average time differ-

ence between two tweets/messages, total number of favourites and re-tweets, and their in-degree and out-degree centrality scores on network of friends and followers were used as features adding to a total of 7 features along with the feature set used in the Topic Modelling experiment (69 LIWC + 89 Topic Modeling words from the Essay Corpus) after observation of the structure of tweets and the previously done linguistic feature experiments. The degree centrality was calculated as of a vertex v, for a given graph $G(V,E)$ with $|V|$ vertices and $|E|$ edges, is defined as: $\{C_D = deg(v)\}$.

The results of all the experiments after 10-fold cross-validation are summarized in Table 6 below.

5 Discussion and Conclusion

The main contributions of this paper are the introduction of a computational Schwartz values model, development of three different corpora annotated with Schwartz' value, and experiments with features for automatic value classification. Table 6 reports that our models outperformed the majority baselines by significant margins of **+5.05, +7.20, +9.83** respectively on the Essay, Twitter and Facebook corpora. From the results it could be inferred that a few Schwartz values such as Self-Direction and Security are relatively difficult to identify, while on the other hand the accuracies for certain value types such as Power and Tradition are persistent and seem to be more salient.

The results also indicate that social media text is difficult for automatic classification, which is obvious from its terse nature. However, it is striking that the social media postings correlate far stronger than the essays with the psychometric data. This is probably since the size of the Twitter data is much larger than someones essay, and since when asked to write something, people become cautious; however, users behave more naturally when communicating in social media, making the data more insightful.

Another major implication from the experiments is that popular text classification features such as n-grams and topic-modelling were not performing well in this domain, indicating that this

---

[9]See for Fine-Gained Speech-Act classes http://compprag.christopherpotts.net/swda.html.

Values	Classifier	Achievement			Benevolence			Conformity			Hedonism			Power		
		SMO	LR	RF	SMO	LR	RF	SMO	LR	RF	SMO	LR	RF	SMO	LR	RF
LIWC	Essay	**65.84**	65.06	64.93	56.06	55.67	**59.58**	**64.01**	61.40	63.49	58.02	**59.20**	54.11	58.80	**59.32**	57.50
+n-grams		57.50	62.71	**65.84**	55.54	53.19	**58.80**	56.45	61.54	**64.80**	58.28	**58.41**	58.02	53.46	**59.71**	58.41
+Topic		58.54	64.15	**65.32**	54.37	53.46	**59.06**	60.63	62.32	**63.75**	**58.80**	58.41	58.28	**58.15**	57.76	56.71
+Lexica		**68.00**	68.00	60.00	**67.00**	65.00	59.00	**75.00**	71.00	63.00	**69.00**	65.00	54.00	**69.00**	67.00	60.00
+Speech-Act		**68.00**	66.80	60.30	**69.00**	67.00	59.00	**71.00**	67.00	59.00	**68.00**	67.00	60.00	**70.00**	67.00	58.00
LIWC	TWT	**80.93**	80.93	80.10	**78.75**	78.75	77.38	73.02	72.48	**77.93**	**77.11**	76.84	76.02	**54.77**	50.68	52.59
LIWC	FB	**85.60**	82.90	81.60	89.10	88.20	**89.90**	**87.50**	86.60	87.50	**85.70**	80.20	80.20	**67.40**	59.20	59.30
+Topic	TWT	74.66	**80.65**	80.65	69.21	**78.20**	77.93	66.76	72.48	**73.02**	71.66	**76.84**	76.57	52.32	**54.77**	51.77
+Topic	FB	79.66	**88.14**	88.14	91.53	**93.22**	93.22	88.14	89.13	**91.53**	83.05	84.75	**86.44**	50.85	**52.54**	50.85
+Lexica	TWT	71.10	**73.70**	69.70	**71.90**	69.90	65.00	67.20	**71.60**	68.00	68.00	**68.60**	60.60	**72.80**	69.80	59.20
+Lexica	FB	**98.20**	86.30	82.60	**93.50**	89.90	89.90	93.90	**96.20**	91.10	**96.80**	81.60	83.90	**91.50**	64.40	56.50
+Non-Linguistic	TWT	74.11	80.38	**80.93**	68.40	**78.47**	77.38	66.49	72.48	**74.11**	70.30	76.30	**76.57**	54.22	**55.59**	54.22
+Speech-Act	TWT	**81.10**	76.40	68.00	**81.00**	73.00	66.00	**75.00**	66.00	66.00	**74.00**	64.00	63.00	**82.00**	75.00	63.00
+Speech-Act	FB	**98.20**	84.50	84.50	**95.90**	89.60	89.60	**93.70**	93.70	90.80	**98.20**	86.60	83.40	**91.20**	66.70	70.30

Values	Classifier	Security			Self-Direction			Stimulation			Tradition			Universalism			Average
		SMO	LR	RF	SMO	LR	RF	SMO	LR	RF	SMO	LR	RF	SMO	LR	RF	
LIWC	Essay	53.06	55.02	**56.06**	**60.89**	59.84	58.54	56.58	**56.98**	56.45	64.28	**65.97**	64.02	65.58	**65.71**	65.32	**61.36**
+n-grams		**56.84**	56.45	56.71	56.06	**58.54**	58.41	56.06	56.67	**56.71**	58.67	**65.06**	64.28	58.28	65.45	**65.84**	**61.05**
+Topic		**56.45**	55.41	54.11	58.67	58.41	**60.76**	56.45	**59.58**	53.59	61.15	**66.10**	66.10	62.45	**65.97**	65.32	**61.40**
+Lexica		**68.00**	66.00	58.00	**73.00**	68.00	62.00	**71.00**	69.00	56.00	**69.00**	65.00	56.00	**71.00**	67.00	62.00	**70.00**
Speech-Act		**73.00**	69.00	58.00	**69.00**	66.00	55.00	**75.00**	71.00	63.00	**74.00**	70.00	62.00	**72.80**	68.30	61.50	**71.15**(+5.05)
LIWC	TWT	**76.29**	75.75	74.11	**83.38**	83.38	75.20	**73.57**	72.48	70.84	**58.04**	55.31	55.86	**82.02**	81.47	80.65	**74.28**
LIWC	FB	**97.50**	97.50	97.50	**85.00**	84.20	83.00	**83.90**	82.80	80.20	**68.60**	59.20	62.00	89.30	**91.00**	88.20	**84.21**
+Topic	TWT	70.57	74.93	**75.48**	76.84	**83.38**	83.38	64.12	**72.47**	71.66	52.04	53.95	**59.67**	74.93	**81.47**	81.20	**73.70**
+Topic	FB	93.22	**98.30**	98.30	86.44	84.75	**89.83**	81.36	84.75	**86.44**	62.71	**74.58**	71.19	89.83	**94.91**	93.22	**85.71**
+Lexica	TWT	70.60	**74.30**	69.50	75.60	74.40	**76.60**	**68.80**	68.60	68.30	**73.90**	69.50	62.30	78.00	**82.20**	76.30	**73.38**
+Lexica	FB	**97.50**	97.50	97.50	**91.60**	82.40	85.00	**92.80**	83.90	83.90	**84.60**	75.10	78.90	90.70	**92.40**	91.60	**93.51**
+Non-Linguistic	TWT	71.18	74.66	**75.20**	76.57	**83.38**	83.38	65.58	**73.57**	71.66	52.59	53.41	**55.86**	74.39	81.74	**82.02**	**73.57**
+Speech-Act	TWT	78.00	**80.00**	69.00	**78.00**	76.00	75.00	**73.00**	66.00	68.00	**80.00**	71.00	63.00	**89.00**	81.10	77.00	**80.00**(+7.20)
+Speech-Act	FB	**97.90**	97.40	97.40	**93.90**	83.60	84.50	**96.30**	85.20	83.94	**91.10**	71.30	78.20	89.50	91.30	**92.20**	**94.50**(+9.83)

Table 6: Automatic Schwartz value classification (accuracy) on the Essay, Twitter and Facebook corpora. Details of feature ablation and class wise performance.

is not yet another text classification problem, but that rather further deeper psycholinguistic analysis is required to find out hidden clues and the nature of language vs ethical practices. Here, it is worth noting the research by Pennebaker (2011), which indicates that, surprisingly, non-content words like pronoun, prepositions, particles, and even symbols are more salient indicators of our personality.

For the machine learners, closer analysis revealed that SMO's performance was somehow irregular and random, which might be an indication of over-fitting. For example, the performance for some Schwartz values greatly decreased when adding n-grams as new features with LIWC, whereas some values showed the opposite behaviour, implying that each value type has its own set of distinct clues, but also high overlap. On the other hand, the performance of the Random Forests classifier increased when the number of features was increased, resulting in a larger forest and hence for most value types it performed better than the other two classifiers with less over-fitting.

A major limitation of the work is that the collected social network corpus is skewed. Reports of psychological analysis on any community always depend on how the target population is chosen. It is absolutely impossible to get precisely balanced data from any real community. For example, it is rather impossible to have 150+ absolute power oriented people in a corpus of size 367 users data. The only solution to this problem is having more data, which we currently are collecting.

The data will be publicly released to the research community soon. We are also very keen on the applied side of this kind of models. Presently we are analysing the community detection problem in social media in relation to values. Another interesting application could be comparative societal analysis between the Eastern and Western regions of the world. Relations among personality and ethics could also be explored.

Acknowledgments

References

Bradley R. Agle and Craig B. Caldwell. 1999. Understanding research on values in business a level of analysis framework. *Business & Society*, 38(3):326–387.

D. Scott Appling, Erica J. Briscoe, Heather Hayes, and Rudolph L. Mappus. 2013. Towards automated personality identification using speech acts. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Antonio Argandoña. 2003. Fostering values in organizations. *Journal of Business Ethics*, 45(1–2):15–28.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Mitja D. Back, Juliane M. Stopfer, Simine Vazire, Sam Gaddis, Stefan C. Schmukle, Boris Egloff, and Samuel D. Gosling. 2010. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21:372–374.

Irénée-Jules Bienaymé. 1853. *Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés*. Imprimerie de Mallet-Bachelier.

Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. The workshop on computational personality recognition 2013. In *Proceedings of the AAAI*, pages 2–5. AAAI.

Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. 2014. The workshop on computational personality recognition 2014. In *Proceedings of the ACM International Conference on Multimedia*, pages 1245–1246. ACM.

C. Joseph Clawson and Donald E. Vinson. 1978. Human values: A historical and interdisciplinary analysis. *Advances in Consumer Research*, 5(1).

Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47. Association for Computational Linguistics.

Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 253–262, New York, NY, USA. ACM.

Lewis R. Goldberg. 1990. An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216.

Jefferson Henrique. 2015. https://github.com/Jefferson-Henrique/GetOldTweets-java.

Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. 1991. *Cultures and organizations: Software of the mind*. McGraw-Hill, 2 edition.

Jacqueline N. Hood. 2003. The relationship of leadership style and ceo values to ethical practices in organizations. *Journal of Business Ethics*, 43(4):263–273.

Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard Hovy. 2013. Learning whom to trust with MACE. In *Proceedings of NAACL-HLT 2013*, pages 1120–1130.

Lynn R. Kahle, Sharon E. Beatty, and Pamela Homer. 1986. Alternative measurement approaches to consumer values: The list of values (LOV) and values and life style (VALS). *Journal of consumer research*, pages 405–409.

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1997–2007, Doha, Qatar, October. Association for Computational Linguistics.

Dejan Markovikj, Sonja Gievska, Michal Kosinski, and David Stillwell. 2013. Mining Facebook data for predictive personality modeling. In *Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA*.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.

Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv:1308.6242*.

James W. Pennebaker, Roger J. Booth, Ryan L. Boyd, and Martha E. Francis, 2015. *Linguistic Inquiry and Word Count: LIWC2015*. Pennebaker Conglomerates, Austin, Texas.

James Pennebaker. 2011. *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury Publishing, New York, New York.

Martin Porter. 1980. An algorithm for suffix stripping. *Program*, 14(3):130–137.

Milton Rokeach. 1973. *The nature of human values*, volume 438. Free Press, New York.

Shalom H. Schwartz and Wolfgang Bilsky. 1990. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of personality and social psychology*, 58(5):878.

Shalom H. Schwartz, Gila Melech, Arielle Lehmann, Steven Burgess, Mari Harris, and Vicki Owens. 2001. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of cross-cultural psychology*, 32(5):519–542.

Shalom H. Schwartz. 2012. An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1):11.

Philip J. Stone, Dexter C. Dunphy, and Marshall S. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet Affect: an affective extension of WordNet. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC-2004)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Pafnuty Lvovich Tchébichef. 1867. Des valeurs moyennes (translated into French by N.V. Khanykov). *Journal de Mathématiques Pures et Appliquées*, 12(2):177–184.

Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. 2014. Sensicon: An automatically constructed sensorial lexicon. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1511–1521, Doha, Qatar, October. Association for Computational Linguistics.

Marc T. Tomlinson, David Hinote, and David B. Bracewell. 2013. Predicting conscientiousness through semantic analysis of Facebook posts. In *Proceedings of the Workshop on Computational Personality Recognition*, pages 31–34.

Michael Wilson. 1988. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1):6–10.

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.